SEMESTER END EXAMINATION, MAY 2024

Course Name: - B.Tech **Semester: 6th**

Paper Name: - Data Mining and Warehousing **Paper Code: TCS 063**

Time - 3Hrs + 20 minutes per hour extra time for V.I. & examinees with writer. **Max Marks-70**

समय– 3 घण्टे+ 20 मिनटप्रति घंटे अतिरिक्त–दृष्टिबाधित एवंसहलेखकपरीक्षार्थियों के लिए। अधिकतम अंक–70

Instructions:

The question paper consists of three sections namely A, B, C. All sections are compulsory.
Section A- Each question carries **03** marks. All questions are compulsory.
Section B- Answer any **5 out of 7** given questions in maximum hundred **(150) words**. Each question carries **7 marks**.
Section C- Answer any **2 out of 3** given questions in maximum onehundred fifty **(500) words**. Each question carries **10 marks**.

निर्देश:

- प्रश्न पत्र मे तीन खण्ड अ, ब, व स हैं। सभी खण्ड अनिवार्य हैं।
- खण्ड–अ में प्रत्येक प्रश्न **3** अंक का है। सभी प्रश्न अनिवार्य हैं।
- खण्ड–ब में 7 प्रश्नों में से किन्हीं 5 प्रश्नों के उत्तर अधिकतम **पचास (150) शब्दों** में दें। प्रत्येक प्रश्न 7 अंक का है।
- खण्ड–स मे 3 प्रश्नों में से किसी 2 प्रश्न के उत्तर अधिकतम **एक सौ प्चास (500) शब्दों** में दें। प्रत्येक प्रश्न 10 अंक का है।

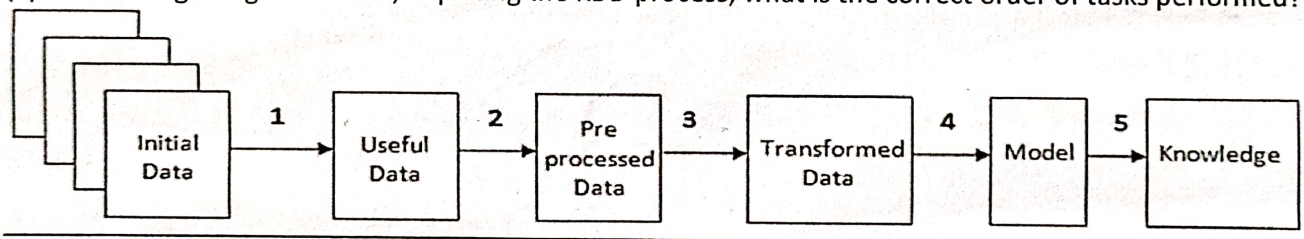
Section – A (खण्ड–अ)

Objective Questions(वस्तुनिष्ठप्रश्न)

1. Answer all the following questions. **5×3=15**
निम्नलिखित सभी प्रश्न अनिवार्य हैं।

(i) Which of the following layer is concerned with producing metadata about the ETL process?
a) Data integration layer b) Applications layer c) Access layer d) None of the above

(ii) In the diagram given below, depicting the KDD process, what is the correct order of tasks performed?



a) Selection, Preprocessing, Transformation, Data mining, Interpretation
b) Selection, Preprocessing, Transformation, Interpretation, Data mining
c) Interpretation, Preprocessing, Transformation, Data mining, Selection
d) Data mining, Preprocessing, Transformation, Selection, Interpretation

(iii) Which of the following is NOT a feature of OLAP?
a) Multidimensional views of data b) Real-time transaction processing
c) Complex calculations d) Time intelligence

(iv) What is the primary goal of outlier analysis?
A) To identify and remove noise from the data
B) To find patterns within the data
C) To identify data points that deviate significantly from the majority of the data
D) To cluster data points into different groups

Scanned with OKEN Scanner

(v) Which of the following uses a set of "if-then" rules for classification?

A) Decision Tree Induction
B) Bayesian Classification
C) Rule-Based Classification
D) Backpropagation

Section – B (खण्ड–ब)

Short Answer Questions (लघुउत्तरीय प्रश्न)

2. Answer any five of the following questions in maximum 150 words. **5×7=35**
नम्नलिखित में से किन्हीं 5 प्रश्नों के उत्तर अधिकतम 150 शब्दों में दें।

(i) What is frequent pattern mining, and what are some commonly used algorithms for it?

(ii) Explain Star, Snow and Fact Constellation Schemas?

(iii) Explain OLAP Operations with an Example?

(iv) Differentiate Between ROLAP, HOLAP and MOLAP Servers?

(v) How is Cluster Analysis applicable in terms of Data Warehousing? And also state some real life application areas of it?

(vi) State the difference between Classification and Prediction?

(vii) How are classification rules generated from a training dataset, and what criteria are used to evaluate their quality? Additionally, how do these criteria impact the performance of rule-based classification systems?

Section – C (खण्ड–स)

Descriptive Questions(विवरणात्मकप्रश्न)

3. Answer any 02 question of the following question in maximum 500 words. **2×10=20**
निम्नलिखित में से किन्हीं 2 प्रश्न के उत्तर अधिकतम 500 शब्दों में दें।

(i) State the top-down approach and bottom-up approach of Data Warehouse Architecture With a diagrammatic representation and also elaborate about both the approaches in brief ?

(ii) By considering this table implement Naïve Bayes Classifier and find Prior probability and Conditional probability of individual attributes

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

And after calculating prior and conditional probability classify new instance into either yes or no class
Condition for new instance : (Outlook=Sunny, Temperature=Cool, Humidity=high, Wind=Storng)

(iii) Explain All steps involve in the process of KDD?

knowledge data, discover operational sqaure external data

Scanned with OKEN Scanner